

A Medical Image Retrieval Framework in Correlation Enhanced Visual Concept Feature Space

Md Mahmudur Rahman, Sameer K. Antani, George R. Thoma
U.S. National Library of Medicine,
National Institutes of Health,
Bethesda, MD, USA
{rahmanmm,santani,gthoma}@mail.nih.gov

Abstract

This paper presents a medical image retrieval framework that uses visual concepts in a feature space employing statistical models built using a probabilistic multi-class support vector machine (SVM). The images are represented using concepts that comprise color and texture patches from local image regions in a multi-dimensional feature space. A major limitation of concept feature representation is that the structural relationship or spatial ordering between concepts are ignored. We present a feature representation scheme as visual concept structure descriptor (VCSD) that overcomes this challenge and captures both the concept frequency similar to a color histogram and the local spatial relationships of the concepts. A probabilistic framework makes the descriptor robust against classification and quantization errors. Evaluation of the proposed image retrieval framework on a biomedical image dataset with different imaging modalities validates its benefits.

1. Introduction

Biomedical Images are commonly stored, retrieved and transmitted in the DICOM (Digital Imaging and Communication in Medicine) format¹ in a Picture Archiving and Communications System (PACS) [2] and image search is on the textual attributes, such as person information, other health meta data, often found in image headers. These attributes are often very brief, however, typically limited to the diagnostic content. It is believed that while improvements in medical image-based diagnoses could be effected through efficient and accurate access to images and related information, their utilization may be limited due to the lack of effective image search methods [1]. Further, search results may be improved by combining text attribute-based

search capability with low-level visual features computed directly on the image content commonly known as Content-Based Image Retrieval (CBIR) [3]. CBIR has the capability to identify visually similar images from a database, however, their relevance may be limited by the “semantic gap”. This gap is introduced due to the limited discriminative power of low-level visual features that are used as descriptors for high-level semantic concepts expressed in an image. In an effort to minimize the semantic gap, some recent approaches have used machine learning on image features extracted from local regions in a partitioned image in a “bag of concepts”-based image representation scheme by treating the features as visual concepts [3]. Such an image representation scheme is based on the “bag of words” representation commonly used in information retrieval from text documents [7]. In this approach, each word is considered independent of all other words and results in loss in document structure. While it has proven effective for text retrieval, it suffers from loss of semantics expressed in a document. This limitation also extends to image retrieval and is further exacerbated because often the correspondence between an image region and local concept is not always always direct [3]. Considering only a single concept per image region while completely ignoring others may lead to two regions matched to different concepts even though they might be very similar or correlated with each other.

This paper presents a spatial correlation-enhanced medical image representation and retrieval framework to address these limitations of the low-level and concept-level feature representation schemes. The organization of the paper is as follows: Section 2 describes the visual concept-based image representation approach. Sections 3 and 4 present a correlation enhanced probabilistic feature representation and structural relationship enhanced feature representation scheme respectively. The experiments and the analysis of the results are presented in Section 5 and Section 6 provides conclusions.

¹<http://www.rsna.org/Technology/DICOM/index.cfm>

2 Image Representation on Local Concept Space

In a heterogeneous collection of medical images, it is possible to identify specific local patches that are perceptually and/or semantically distinguishable, such as homogeneous texture patterns in grey level radiological images, differential color and texture structures in microscopic pathology and dermoscopic images. The variation in these local patches can be effectively modeled by using supervised learning based classification techniques such as the Support Vector Machine (SVM) [8]. In its basic formulation, the SVM is a binary classification method that constructs a decision surface and maximizing the inter-class boundary between the samples. However, a number of methods have been proposed for multi-class classification.

For concept model generation, we utilize a voting-based multi-class SVM known as *one-against-one* or pairwise coupling (PWC) [9]. In developing training samples for this SVM, only local image patches that map to visual concept models are used. A fixed-partition based approach is used at first to divide the entire image space into a $(r \times r)$ grid of non-overlapping regions. Manual selection is applied to limit such patches in the training set to those that have a majority of their area (80%) covered by a single semantic concept. In order to perform the multi-class SVMs training based on the local concept categories, a set of L labels are assigned as $C = \{c_1, \dots, c_i, \dots, c_L\}$, where each $c_i \in C$ characterizes a local concept category. Each patch is labeled with only one local concept category and is represented by a combination of color and texture moment-based features.

Images in the data set are annotated with local concept labels by partitioning each image I_j into an equivalent $r \times r$ grid of l region vectors $\{\mathbf{x}_{1j}, \dots, \mathbf{x}_{kj}, \dots, \mathbf{x}_{lj}\}$, where each $\mathbf{x}_{kj} \in \Re^d$ is a combined color and texture feature vector. For each \mathbf{x}_{kj} , the local concept category probabilities are determined by the prediction of the multi-class SVMs as [9]

$$p_{ikj} = P(y = i \mid \mathbf{x}_{kj}), \quad 1 \leq i \leq L. \quad (1)$$

Based on the probability scores, the category label of x_{kj} is determined as c_m as the label with the maximum probability score. Hence, the entire image is thus represented as a two-dimensional index linked to the concept or localized semantic labels assigned for each region. Based on this encoding scheme, an image I_j can be represented as a vector in a local semantic concept space as

$$\mathbf{f}_j^{\text{Concept}} = [f_{1j}, \dots, f_{ij}, \dots, f_{Lj}]^T \quad (2)$$

where each f_{ij} corresponds to the normalized frequency of a concept $c_i, 1 \leq i \leq L$ in image I_j . However, this representation captures only a coarse distribution of the concepts. It is very sensitive to quantization or classification

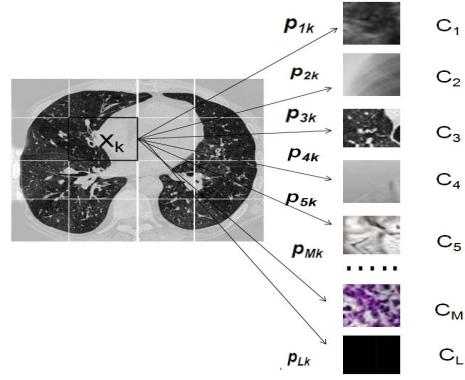


Figure 1. Probabilistic membership scores

errors and ignores correlations and structural relationships among concepts.

3 Probabilistic Feature Representation

The feature vector $\mathbf{f}_j^{\text{concept}}$ can be viewed as a local concept distribution from a probabilistic viewpoint. Given a set of concept categories of length L , each element f_{ij} of $\mathbf{f}_j^{\text{concept}}$ for an image I_j is calculated as $f_{ij} = l_i/l$. It is the probability of a region in the image encoded with label i of the concept $c_i \in C$, and l_i is the total number of regions that map to c_i . According to the total probability theory [10], f_{ij} can be defined as

$$f_{ij} = \sum_{k_j=1}^l P_{i|k_j} P_k = \frac{1}{l} \sum_{k_j=1}^l P_{i|k_j} \quad (3)$$

where P_k is the probability of a region selected from image I_j being the k_j th region, which is $1/l$, and $P_{i|k_j}$ is the conditional probability that the selected k_j th region in I_j maps to the concept c_i . In the context of the concept vector $\mathbf{f}_j^{\text{concept}}$, the value of $P_{i|k_j}$ is 1 if the region k_j is mapped to the c_i concept, or 0 otherwise. Due to the crisp membership value, this feature representation is sensitive to quantization errors.

We present a feature representation scheme based on the observation that there are usually several concepts that are highly similar or correlated to the best matching one for a particular image region. This scheme spreads each region's membership values or confidence scores to all the local concept categories. During the image encoding process, the probabilistic membership values of each region to all concept prototypes are computed for an image I_j . For example, Figure 1 shows a particular region in a segmented image and its probabilistic membership scores to different local concept categories. Based on the probabilistic values of each region, an image I_j is represented as

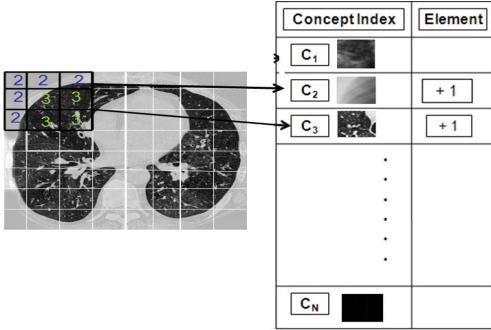


Figure 2. Visual concept structure descriptor

$$\mathbf{f}_j^{\text{PVCV}} = [\hat{f}_{1j} \cdots \hat{f}_{ij} \cdots \hat{f}_{Lj}]^T, \text{ where}$$

$$\hat{f}_{ij} = \sum_{k=1}^l p_{ikj} P_k = \frac{1}{l} \sum_{k=1}^l p_{ikj}; \quad \text{for } i = 1, 2, \dots, L \quad (4)$$

where p_{ikj} is determined based on (1). Here, we consider each of the regions in an image being related to all the concepts via the membership values such that the degree of association of the k_j -th region in I_j to the c_i concept is determined by distributing the membership values to the corresponding index of the vector. In contrast to the simple concept vector $\mathbf{f}^{\text{concept}}$, this vector representation considers not only the similarity of different region vectors from different concepts but also the dissimilarity of those region vectors mapped to the same concepts.

4 Structural Feature Representation

A major limitation of concept feature representation is that the structural relationship or spatial ordering between concepts are ignored. This representation can not distinguish between two images in which a given concept is present in identical numbers but where the structure of the groups of regions having that concept is different. We present a feature representation scheme as *visual concept structure descriptor* (VCSD) that overcomes this challenge and captures both the concept frequency similar to a color histogram and the local spatial relationships of the concepts. Specifically, it is a vector $\mathbf{f}_j^{\text{VCSD}} = [f_{1j}^v \cdots f_{ij}^v \cdots f_{Lj}^v]^T$, where each element f_{ij}^v represents the number of times a visual concept label is present in a windowed neighborhood determined by a small square structuring element. The size of the structuring element is $(b \times b, b < r)$ units. This is illustrated in Figure 2 where an image is partitioned into 64 blocks ($r = 8$). A 9-element ($b = 3$) structuring element enables distinction between images with the same concepts that are in equal proportions on their distribution. The structuring element is moved over the image in an overlapping

fashion and accumulates the visual concept labels. This process is also illustrated in the figure. For each unique concept at a particular position in the image within the structuring element, the corresponding element of the feature vector is incremented. Upon completion, the concept vector is normalized by the number of positions of the structuring element.

5 Experiments and Results

The image collection for experiment comprises of 5000 bio-medical images of 30 manually assigned disjoint global categories, which is a subset of a larger collection of six different data sets used for medical image retrieval task in ImageCLEFmed 2007 [5]. In our collection, the images are classified into three levels as modalities, body parts, orientations or distinct visual observation. For the SVM training, 30 local concept categories, such as tissues of lung or brain of CT or MRI, bone of chest, hand, or knee X-ray, microscopic blood or muscle cells, dark or white background, etc. are manually defined. The training set used for this purpose consist of only 5% images of all global categories of the entire data set. To generate the local patches, each image in the training set is at first partitioned into an 8×8 grid generating 64 non-overlapping regions. Only the regions that conform to at least 80% of a particular concept category are selected and labeled with the corresponding category label. For the SVM training, a 10-fold cross-validation (CV) is conducted to find the best values of tunable parameters $C = 200$ and $\gamma = 0.02$ of the radial basis function (RBF) kernel with a CV accuracy of 81.01%. We utilized the LIBSVM² software package for implementing the multi-class SVM classifier.

For a quantitative evaluation of the retrieval results, we selected all the images in the collection as query images and used *query-by-example (QBE)* as the search method. Figure 3 shows the precision-recall curves based on the Euclidean similarity matching in different feature spaces. The performance was compared to the low-level MPEG-7 based color layout descriptor (CLD) and edge histogram descriptor (EHD) [11]. By analyzing the Figure 3, we can observe that the proposed concept-based feature representation schemes performed much better compared to the low-level MPEG-7 (e.g., CLD and EHD) based features in terms of precision at each recall level. The better performances are expected as the concept features are more semantically oriented that exploits the domain knowledge of the collections at a local level. It is also noticeable that, the performances of both the probabilistic visual concept vector (PVCV) and visual concept structure descriptor (VCSD) increase at a lower recall level (e.g., up

²<http://www.csie.ntu.edu.tw/~cjlin/libsvm>

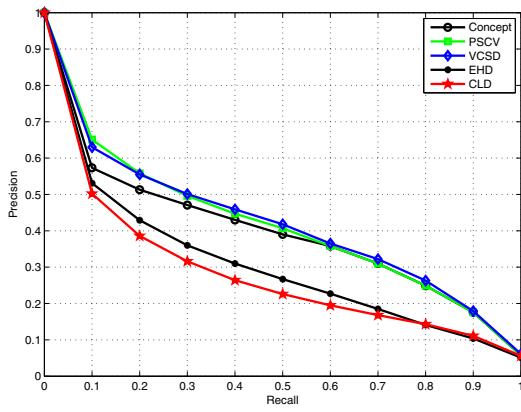


Figure 3. Precision-recall curves in different feature spaces.

to 0.6) when compared to the normalized frequency based feature vector (e.g., Concept). These results are encouraging enough as users are mainly interested to find relevant images in only few retrieved images (e.g., at a low recall level). From the results, we can conjecture that there is always enough correlation and structural relationships between the local concepts, which can be exploited in the feature representation schemes.

6 Conclusions

This paper proposes new techniques for improving accuracy of medical image retrieval by representing image content at an intermediate level local visual concept level. The intermediate level is higher than low-level visual features that are traditionally used and a step closer to the high-level semantics in the image content. A visual concept is defined for local image regions and an image may comprise of several concepts. The feature space is enhanced by exploiting the correlations and structural relationships among the these visual concepts. Using SVM-based training, the proposed image representation schemes realize semantic abstraction via prior learning when compared to the representations based on the low-level features. Experimental results validate the hypothesis and shows that the proposed representation schemes improve overall retrieval accuracy.

Acknowledgment

This research is supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and Lister Hill National Center for Biomedical Communications (LHNCBC). We

also thank the imageCLEFmed [5] organizers for making the database available for the experiments.

References

- [1] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, “A Review of Content-Based Image Retrieval Systems in Medical Applications Clinical Benefits and Future Directions,” *International Journal of Medical Informatics*, vol. 73, pp. 1–23, 2004.
- [2] R.H. Choplin, J.M. Boehme, and C.D. Maynard, “Picture archiving and communication systems: an overview,” *RadioGraphics*, vol. 12, pp. 127–129, 1992.
- [3] Y. Liua, D. Zhang, G. Lu, and W.Y. Ma, “A survey of content-based image retrieval with high-level semantics,” *Pattern Recognition*, vol. 40, pp. 262–282, 2007.
- [4] W. Hsu, S. Antani, L.R. Long, L. Neve, and G.R. Thoma, “SPIRS: a Web-based Image Retrieval System For Large Biomedical Databases,” *International Journal of Medical Informatics*, vol. 78, pp. 13–24, 2008.
- [5] H. Müller, T. Deselaers, E. Kim, C. Kalpathy, D. Jayashree, M. Thomas, P. Clough, and W. Hersh, “Overview of the ImageCLEFmed 2007 Medical Retrieval and Annotation Tasks,” *8th Workshop of the Cross-Language Evaluation Forum (CLEF 2007), Proceedings of LNCS*, vol. 5152, 2008.
- [6] T.M. Lehmann, B.B. Wein, J. Dahmen, J. Bredno, F. Vogelsang, and M. Kohnen, “Content-based image retrieval in medical applications-A novel multi-step approach,” *Proc SPIE*, vol. 3972, pp. 312–320, 2000.
- [7] R.B. Yates and B.R. Neto, *Modern Information Retrieval*, 1st ed., Addison Wesley, 1999.
- [8] V. Vapnik, *Statistical Learning Theory*, New York, NY, Wiley; 1998.
- [9] T.F. Wu, C.J. Lin, and R.C. Weng, “Probability Estimates for Multi-class Classification by Pairwise Coupling,” *Journal of Machine Learning Research*, vol. 5, pp. 975–1005, 2004.
- [10] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Boston, 2nd edition: Academic Press; 1990.
- [11] S.F. Chang, T. Sikora, and A. Puri, “Overview of the MPEG-7 standard,” *IEEE Trans Circ Syst Video Technology*, vol. 11, pp. 688–695, 2001.